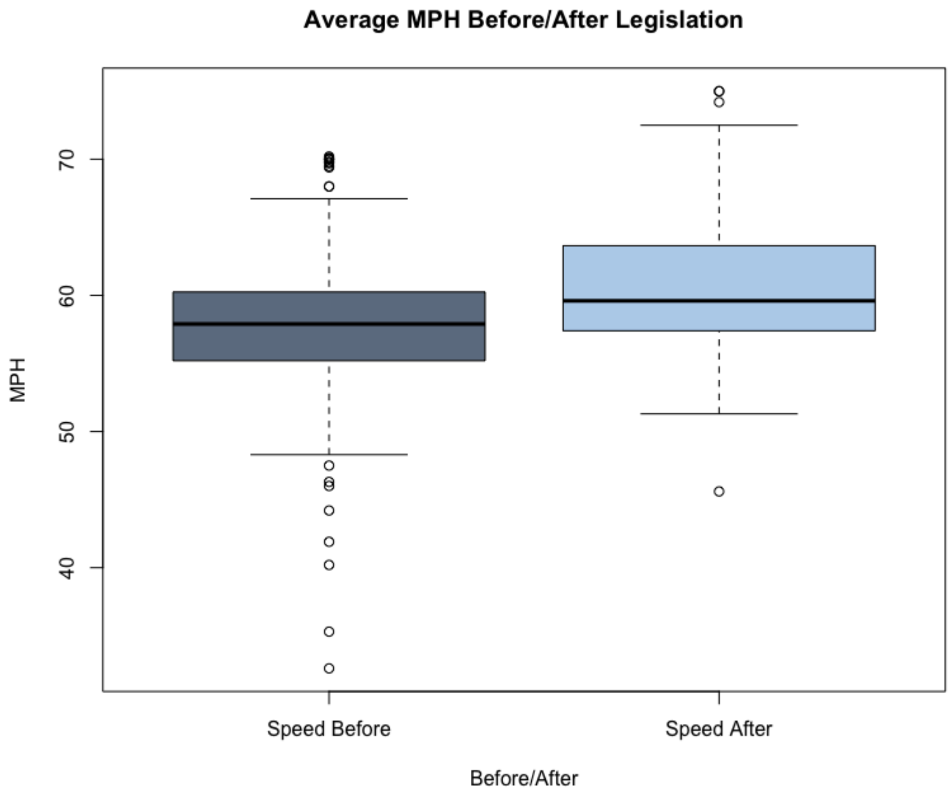
CS5590-0025
Homework #2

Thomas Jones
08206472

1. Generate boxplots for vehicular before and after speeds data. Discuss and compare the results.



The box plot above shows the speeds before and after the repeal of the national maximum speed limit on U.S. roads. The bar plot clearly shows a general increase in average speed, as would be expected with a repeal of a national maximum.

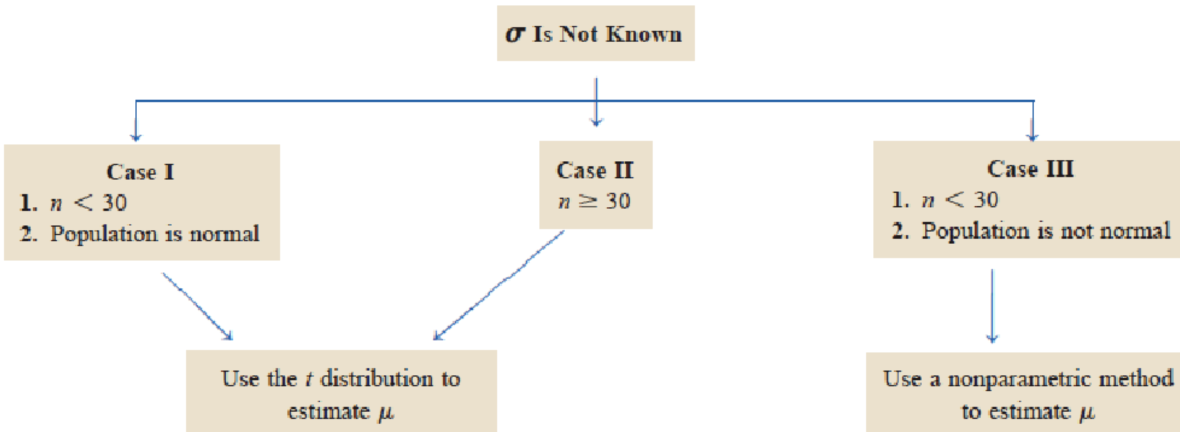
Timeframe	Mean	Median	SD	Skewness	Kurtosis
Before	57.77251	57.9	4.301547	-0.4673015	2.81653636
After	60.67951	59.6	4.520668	0.6042475	-0.03205487

The summary statistics for the data are shown above. As shown in the box plot the average speed increased after the repeal. The data also shows that after the repeal the skew of speeds shifted from a negative to positive skew indicating that the majority of speeds shifted from above the average to below the average, i.e. while speeds generally increased the trend was slower compared to the average. The kurtosis indicates that after the repeal the data is closer to a normal distribution as opposed to prior to the repeal where before the repeal the distribution was sharp with fatter tails. This is visually confirmed in the box plot which shows a higher number of outliers.

Thomas Jones
08206472

2. Generate 99% confidence intervals for mean vehicular before-speed data, assuming the population variance is unknown. Explain each step and interpret the results.

Following the flow chart:



The length of the speed data before repeal is 764 so a t-distribution is used to estimate the confidence interval.

The $(1-\alpha)100\%$ confidence interval for μ

$$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}} \quad \text{Where,} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

From the summary table we know that the sample mean is 57.77 and sample standard deviation is 4.30. The standard error is then 0.16.

We use a value of 0.995 since, i.e. 1.0 - (1.0-0.99)/2. This yields a critical value of 2.58.

Finally, plugging this in the 99% confidence interval is [57.37, 58.17].

Given the data provided, the probability that the true mean of the speed data before the repeal was between 57.37 and 58.17 is 99%.

Thomas Jones
08206472

3. Generate 90% confidence intervals for the variance of after-speed data. Explain each step and interpret the results.

Calculating confidence intervals for variance is done with the following formula:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

For the after repeal data there are a number of NA values. For this data n = 571. The sample variance is 20.44.

The lower and upper chi-squared values are 626.65 and 626.65 respectively.

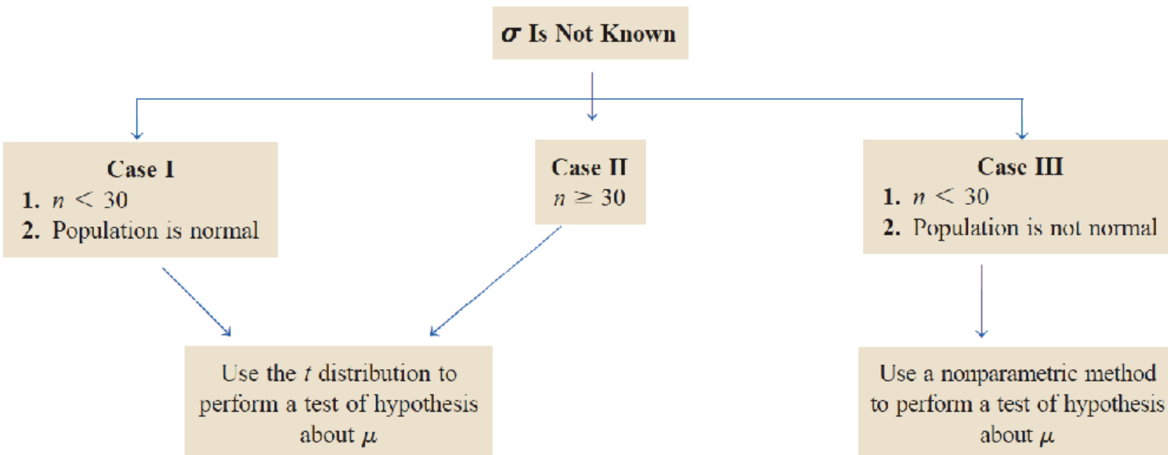
Plugging this into the equation above the 90% confidence interface are then [18.59, 22.59].

Given the data provided, there is a 90% probability that the variance of the speed data after the repeal is between 18.59 and 22.59.

4. Test whether the mean speed is equal to 65 mph after repealing the speed limit at the α=1% significance level. Write each step of the hypothesis test and interpret the results.

The null hypothesis is that the mean speed is not 65 mph after repealing the speed limit.

The alternative hypothesis is that the mean speed is 65 mpg.



Since the true variance is not known and n >= 30 the t-distribution is used to perform the test. The distribution is normal and the confidence level is 0.01

The test statistic t is found by,

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

The resulting statistic is with sample mean 60.68, sample std = 4.52 and sample n = 517 is -22.84.

The resulting p-value is 1.0. Since we are using an equality test (or two-tailed test) we use 2*p-value or 2.0. Since 1.0 < 0.01 we cannot reject the null hypothesis and the actual mean speed may be 65 mph.

Thomas Jones
08206472

5. Test whether the variance of after-speed data is less than 18 mph^2 at the α=5% significance level. Write each step of the hypothesis test and interpret the results.

The null hypothesis is that the variance of the after-speed data is greater than or equal to than 18 mph^2.

The alternative hypothesis is that the variance is less than 18 mph^2, i.e. greater than.

The significance level is 0.05. The distribution to be used is chi-squared.

The test statistic is given as,

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Given n = 517, the sample variance = 20.44, this gives the test statistic value = 647.15. Since we are considering “less than” this is a left tailed test.

The critical value is calculated as 515.62 and since the test statistic is not less than the critical value, the null hypothesis fails to be rejected, i.e. we cannot say if the after speed data is not less than 18 mph^2.

6. **Test that the vehicular before-speed variance is less than after-speed at the α=10% significance level. Write each step of the hypothesis test and interpret the results.

The null hypothesis is that the before speed variance is greater than or equal to the after speed variance.

The alternate hypothesis is that the before speed variance is less than the after speed variance.

To determine this the ratio of the variances is calculated as our F statistic which is 0.91. Next, the F critical value is calculated for our 0.1 significance level. This value is 0.90.

To determine if the null hypothesis is rejected the F statistic and critical values are compared and if the statistic is less than the critical value then the hypothesis is rejected. A left-tailed test will be used.

For the datasets given, 0.91 > 0.90 so we reject the null hypothesis that the before variance is less than the after variance.

7. Test that the vehicular after-speed mean is greater than before-speed at the α=5% significance level. Write each step of the hypothesis test and interpret the results.

The null hypothesis is that the after speed mean is not greater than the before speed mean.

The alternative hypothesis is that the after speed mean is greater than the before speed mean.

The test will be right tailed.

To verify this the degree of freedom must be calculated from,

$$df = \frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\frac{\left(\frac{{s_1}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{{s_2}^2}{n_2}\right)^2}{n_2 - 1}}$$

For the samples this comes to 1194.

Then the value of t,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

The value of t for samples is -11.87.

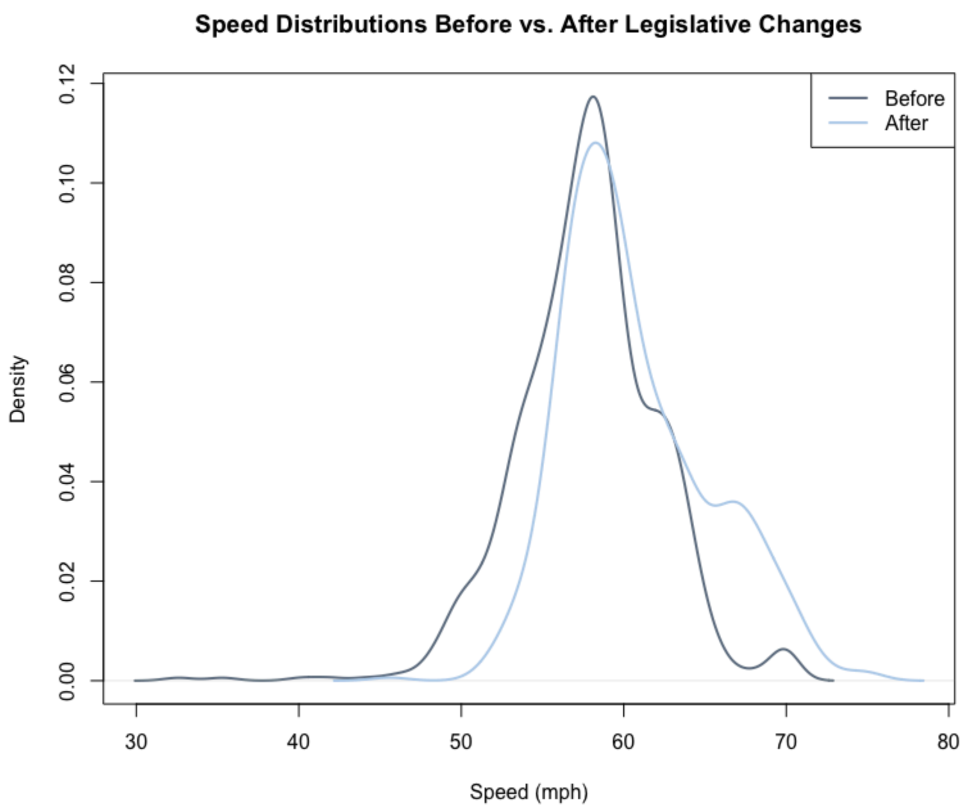
This yields a value for p of 0.00 which is less than 0.05 so we reject the null hypothesis and say that the mean speed after the repeal is greater than the mean speed before the repeal.

Thomas Jones
08206472

8. Use a Mann-Whitney-Wilcoxon test at the α=5% significance level to assess whether the distributions of speeds before and after are equal. Draw density plots using before and after speed data. Interpret the results based on both the test and drawing.

The null hypothesis is that the two sample distributions are drawn from the same population. The alternative hypothesis is that the two sample distributions are drawn from two different populations.

Running the Mann-Whitney-Wilcoxon test the resulting p value is 0.00 so we reject null hypothesis and say that the two distributions come from different populations.



The above distribution changes, while both normal and sharing some of the same mean values have different characteristic shapes. The Mann-Whitney-Wilcoxon test is non-parametric, i.e. it does not consider statistical data. The results also indicate that there is strong evidence that repeal of speed limits had an effect on the speed of drivers, i.e. that the changes in driver behavior across the samples was not due to chance. This is clearly indicated in the density plot as well as the upward slope of the distribution starts at a higher speed for the after data. There is also a fat tail on the right side of the distribution indicating generally more drivers going faster than 65 mph.